Learning under Noisy Labels, Spurious Points, and Diverse Structures: TS40K, a 3D Point Cloud Dataset of Rural Terrain and Electrical Transmission Systems

Diogo Lavado
NOVA SST & UniMi
d.lavado@campus.fct.unl.pt

Ricardo Santos
EDP - Labelec
ricardovieira.santos@edp.com

André Coelho
EDP - Labelec
andre.coelho@edp.com

João Santos
CNET - Centre New Energy
joao.passagemsantos@edp.pt

Alessandra Micheletti
UniMi
alessandra.micheletti@unimi.it

Cláudia Soares
NOVA SST
claudia.soares@fct.unl.pt

Abstract

Research in 3D scene understanding, particularly in autonomous driving and indoor segmentation, has made significant strides. However, most available datasets focus on urban settings. We introduce TS40K, a 3D point cloud dataset spanning 40,000 km of electrical transmission systems in rural terrain, addressing power-grid inspections to prevent outages, damages, and fires. TS40K offers high point density and no occlusion, presenting challenges like noisy labels, diverse structures, and sensor noise causing spurious points. We evaluate state-of-the-art methods on 3D semantic segmentation and object detection, revealing limitations in power grid inspection. TS40K invites further research to tackle these challenges. Resources available in: https://github.com/dlavado/TS40K

1. Introduction

Accurate 3D scene understanding is essential for real-world applications, yet most existing datasets focus on urban scenarios like autonomous driving [36] and indoor scene segmentation [1, 10]. Current models, such as JS3C-Net [62] and RangeFormer [26], are designed primarily for urban settings, limiting the exploration of diverse environments. To address this gap, we introduce **TS40K**, a large-scale 3D point cloud dataset spanning over 40,000 kilometers of electrical transmission systems in rural terrain. TS40K enables advancements in both 3D semantic segmentation, with per-point annotations across five semantic classes, and 3D object detection, focusing on critical elements of power-grid infrastructure.

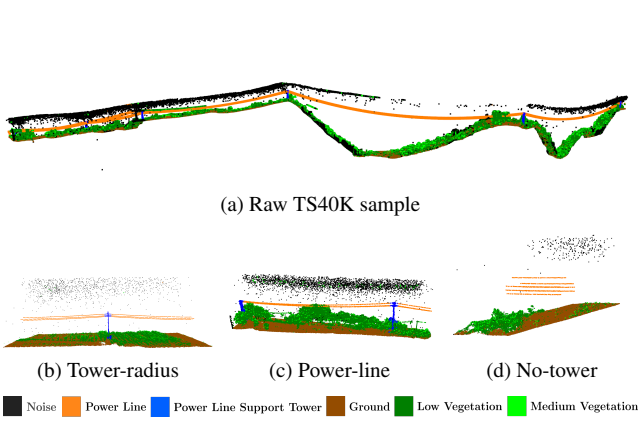
3D scene understanding in electrical transmission systems enhances inspection efficiency by detecting defects, identifying collision risks, and reducing the risk of power outages, grid damage, and forest fires. Drones equipped with LiDAR sensors are increasingly used for power-grid scans, but these scans are processed manually by maintenance personnel, a time-consuming task involving the annotation of each 3D point. With aging infrastructure and increasing demand for reliability, the need for more efficient inspection methods has grown, especially given the significant economic and environmental impacts of grid failures and forest fires [42, 52].

TS40K introduces several key challenges to the 3D scene understanding community: **(1) High-density noise.** LiDAR scans are often affected by noise, which can obscure power-grid elements, particularly under adverse weather conditions. **(2) Inspection-based annotations.** The dataset includes annotations designed for maintenance, not machine learning models, leading to mislabeled points and offering a more realistic evaluation of AI methods. **(3) Objects with diverse structures and low point density.** The dataset presents extreme class imbalance, with ground and vegetation dominating, while power-grid elements are underrepresented (1.4% of the data). Our contributions are threefold: **(1) Novel Dataset:** TS40K is the first publicly available 3D point cloud dataset focused on rural power-grid systems. **(2) Method Evaluation:** We evaluate state-of-the-art methods for 3D semantic segmentation and object detection on this dataset. **(3) Challenges:** TS40K addresses noisy labels, spurious points, and diverse structures, making it a valuable resource for advancing 3D scene understanding in complex, real-world environments.

2. Related Work

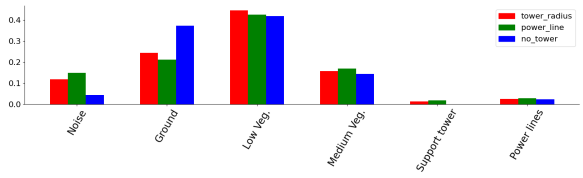
2.1. Contemporary 3D Datasets

The progress in 3D scene understanding techniques is largely substantiated by high-quality, large-scale, and densely annotated datasets [55]. 3D benchmarks currently

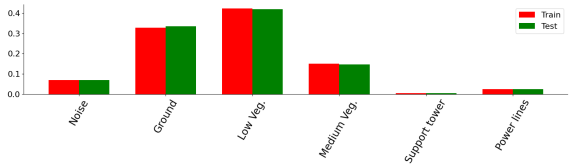


(a) Raw TS40K sample

(b) Tower-radius (c) Power-line (d) No-tower



(e) Sample type density



(f) Overall train/test class density

Figure 1. The TS40K dataset is derived from raw 3D scans illustrated in Figure 1a and processed into three different sample types: *(1) Tower-radius* focuses on the towers that support power lines and its environment (Fig. 1b). *(2) Power-line* samples have power lines as their main focus in the 3D scenes (Fig. 1c). *(3) No-tower* samples represent rural terrain where the transmission system is located, excluding supporting towers but potentially including power lines (Fig. 1d). In Figures 1e and 1f, we showcase the semantic class densities of the TS40K dataset. Figure 1e illustrates the class density for each of the sample types and Figure 1f shows the overall class density in the TS40K train and test sets.

available can be broadly categorized into three groups: **(1) Outdoor urban datasets.** Most of these datasets focus on self-driving applications and complex urban structures, namely, the KITTI benchmark [2, 15, 16], KAIST [23], Lyft dataset [25], Cityscapes 3D [14], nuScenes [3], Waymo dataset [50], SensatUrban [20] and the ONCE dataset [35]. **(2) Indoor 3D scans.** These benchmarks include heavily detailed indoor environments, such as NYU3D [47], SUN RGB-D [48], SceneNN [22], S3DIS [1] and ScanNet [10]. **(3) 3D object representations.** These datasets include every-day objects for classification or part-segmentation, for instance, MeshSeg [6], ShapeNet [4], PartNet [39], and ScanObjectNN [57]. Conversely to these benchmarks, there are no publicly available datasets that focus on non-urban areas and that follow power-grid systems. For example, 3D datasets that cover rural or forest terrain are Forest3D [56], GTASynth [9] and NEON [37], whereas DALES [58] includes power-grid elements in urban areas. Forest3D [56] focuses solely on 3D representations of trees and in their instance segmentation. GTASynth [9] is a synthetic dataset with of non-urban environments with 3D point cloud properties similar to self-driving datasets, specifically low point density, object occlusion and captured from a vehicle point of view. On the other hand, NEON [37] captures airborne LiDAR tree data from a birds-eye view perspective to predict tree crown dimensions. Lastly, DALES [58] focuses on urban scenes with high voltage towers, whereas TS40K features medium and small voltage towers on rural areas, which are more diverse in shape and harder to discern from the environment. In contrast to these datasets, the TS40K dataset focuses on electrical power-grid systems and their environment, which include low and medium vegetation, trees, and highly irregular terrains. Additionally, TS40K is composed of medium and large point clouds that feature high point density, no object occlusion, and homogeneous object density. Conversely, urban datasets usually contain sparse point clouds with occluded objects, limiting performance, while indoor and 3D object datasets tend to focus on small point clouds within closed environments. A comprehensive comparison between prominent publicly available benchmarks and TS40K can be found in Table 1.

Table 1. Key attributes of various 3D datasets prominent in 3D scene understanding. Entries include the dataset's publication year, viewing perspective, realism (real or simulated), total point count, number of semantic classes and the number of annotated classes in brackets, bounding box cardinality, and inclusion of RGB information. Forest datasets were excluded from this overview due to the absence of crucial information, such as the number of points or classes and the number of bounding boxes.

3D Dataset	Year	View	Real/Simulated	#Points	#Classes	#Bounding Boxes	RGB
KITTI [15, 16]	2012	single vehicle	Real	1799M	3 (8)	200K	No
KAIST [23]	2018	single vehicle	Real	-	3 (3)	103K	Yes
SemanticKITTI [2]	2019	single vehicle	Real	4549M	25 (28)	-	No
Lyft [25]	2019	single vehicle	Real	-	9 (9)	1.3M	No
Waymo Open [50]	2019	single vehicle	Real	-	4 (4)	12M	No
nuScenes [3]	2019	single vehicle	Real	1170M	23 (23)	1.4M	No
SensatUrban [20]	2020	UAV	Real	2847M	13 (31)	-	Yes
ScanNet [10]	2017	RGB-D	Real	242M	20 (20)	-	Yes
S3DIS [1]	2017	RGB-D	Real	273M	13 (13)	-	Yes
ShapeNet [4]	2015	-	Simulated	-	55 (55)	-	No
PartNet [39]	2019	-	Simulated	-	24 (24)	-	No
DALES [58]	2021	UAV	Real	505M	8 (8)	-	No
TS40K (Ours)	2024	UAV	Real	2595M	5 (22)	36K	No

2.2. 3D Semantic Segmentation

Semantic segmentation, operating at the scene level, seeks to partition a 3D point cloud into subsets based on the semantic meanings attributed to individual points. Semantic segmentation methodologies can be broadly categorized

into four paradigms [18]: **(1) projection-based.** [29, 34, 49, 66] These methods employ established 2D CNN frameworks to learn 3D semantics. However, projecting point clouds onto 2D images introduces the risk of losing critical geometric information. **(2) discretization-based.** [7, 30, 38, 69, 71] These models leverage 3D CNN architectures. While effective, these techniques often encounter scalability challenges due to significant computational and memory requirements. **(3) point-based.** [21, 26, 27, 33, 40, 41, 54, 59] These methods adopt downsampling (set-abstraction) and upsampling (feature-propagation) techniques directly on point clouds, in the form of and. In contrast to voxel and projection-based methods, point-based architectures preserve the semantics for each individual 3D point and achieve state-of-the-art performance in most prominent datasets. Lastly, **(4) hybrid methods.** [11, 19, 24, 51] These methods take advantage of the three techniques described above and fuse the feature extraction knowledge from each channel to achieve better domain understanding of 3D scenes. In 3D semantic segmentation, new methods emerge regularly, challenging established benchmarks, particularly those tailored for autonomous driving. These benchmarks dominate the field, shaping research priorities towards improving performance in such scenarios. To diversify the evaluation landscape, our dataset addresses the specific challenges of power-grid inspection, a critical task for preventing power outages and forest fires.

2.3. 3D Object Detection

3D object detection aims to predict bounding boxes of 3D objects in 3D scenes. A general formula of 3D object detection can be represented as $B = f_{det}(\mathcal{I})$, where $B = \{B_1, \ldots, B_N\}$ is a set of N 3D objects in a scene, f_{det} is a 3D object detection model, and $\mathcal{I}$ is one or more sensory inputs. A 3D object B_i is represented as a 3D cuboid, including its parameters as $B_i = [x_c, y_c, z_c, l, w, h, \theta, \text{class}]$, where (x_c, y_c, z_c) is the 3D center coordinate of the cuboid, l, w, h represent its length, width, and height, respectively. θ is the heading angle of the cuboid on the ground plane, and class denotes the category of the 3D object [36]. Our dataset follows this definition to define bounding boxes. 3D object detection methodologies can be broadly categorized in two groups [18]: **(1) Region proposal-based methods.** [44, 45, 65, 67, 71] These techniques initially generate multiple potential regions, often referred to as proposals, that encompass objects. Subsequently, they extract features specific from each region to filter/refine region proposals and ascertain their category label. **(2) Single shot methods.** [12, 31, 61, 64, 68] These approaches directly forecast class probabilities and perform 3D bounding box regression using a single-stage network. Dispensing the need for a region proposal stage and subsequent post-processing, they show accelerated processing speed. As indicated in Table 1,

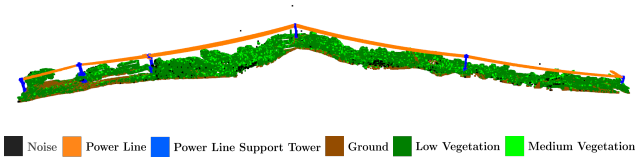


Figure 2. Example of TS40K raw 3D point clouds.

Table 2. Annotated classes in the TS40K dataset and their distribution for power-grid inspection. Ground and road surfaces constitute the majority of the dataset (63%), whereas the power-grid only constitute 1.43% of the 3D points.

Label	Class	Density(%)	Label	Class	Density(%)
0	Created	0	11	Road surface	**44.752**
1	Unclassified	0.571	12	Overlap points	0.529
2	Ground	**23.403**	13	Medium Reliability	0
3	Low vegetation	**18.758**	14	Low Reliability	0
4	Medium vegetation	0.241	15	Power line support tower	**0.519**
5	Natural obstacle	1.069	16	Main power line	**0.907**
6	Human structures	0	17	Other power line	**0.002**
7	Low point	0.362	18	Fiber optic cable	0
8	Model key points	0	19	Not rated object to be consider	8.205
9	Water	0	20	Not rated object to be ignored	0
10	Rail	0.681	21	Incidents	0

3D bounding box annotations for objects within 3D scenes are predominantly available in autonomous driving benchmarks. Our TS40K dataset contributes to broadening the scope of 3D object detection benchmarks by introducing a focus on power-grid elements in a rural setting. Notably, we offer 3D bounding boxes for power lines, their supporting towers and medium vegetation. This inclusion enhances the ability of inspectors to assess the risk of grid contact with its surroundings.

3. The TS40K Dataset

3D Point Cloud Collection. The TS40K dataset utilizes 3D point cloud data obtained from unmanned aerial vehicles (UAVs) conducting scans of electrical transmission systems. Notably, the use of UAVs results in capturing data from a birds-eye view perspective, this leads to good data characteristics for learning models, such as high point density, absence of object occlusion, and homogeneous object density. Our UAV operates with a 70.4º FOV and captures data at a rate of 240K shots per second, generating 720K points per second at a density of 170 points per square-meter. At a typical flying altitude of 100 meters, it covers a swath width of 140 meters. The system achieves a precision of 2.5cm and an accuracy of 3.0cm, making it suitable for detailed terrain mapping. Data is recorded onto USB flash drives for post-flight processing, and optional real-time monitoring is available via radio-modem transmission. The system merges LiDAR and GNSS data into geo-referenced point clouds, which can then be used to generate high-quality samples for further analysis. Our dataset comprises over 500 processed and annotated scans, each with 80 km on average, by maintenance personnel and encompasses 40,000 km of diverse land strips dedicated to the transmission system (Figure 2). Each 3D point is manually

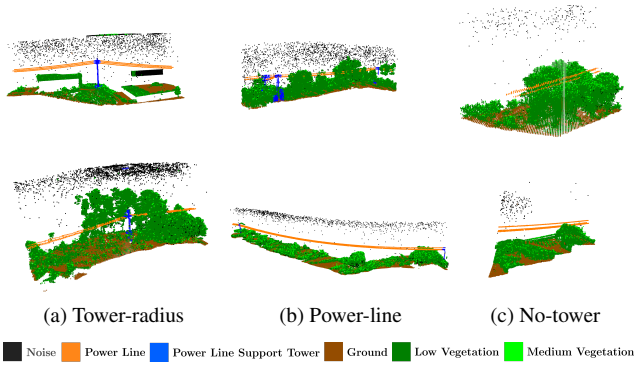


Figure 3. The raw TS40K land strips are partitioned into three sample types: *Tower-radius*: Encompasses areas around power-line supporting towers (Fig. 3a); *Power-line*: Focuses on power lines between towers (Fig. 3b); and *No-tower*: Represents rural areas without towers but potentially including power lines(Fig. 3c). This categorization ensures safety and addresses data imbalance.

Table 3. Mapping of annotated labels to semantic classes for power-grid inspection. The semantic classes were developed in collaboration with maintenance personnel to highlight key scene elements for inspections. '—' describes annotated classes that do not have a mapping due to absence in the original data.

Original Label	Annotated Class	Semantic Class	Original Label	Annotated Class	Semantic Class
0	Created	—	11	Road surface	Ground
1	Unclassified	Noise	12	Overlap points	Low Vegetation
2	Ground	Ground	13	Medium reliability	—
3	Low vegetation	Low Vegetation	14	Low reliability	—
4	Medium vegetation	Medium Vegetation	15	Power line support tower	Power line support tower
5	Natural obstacle	Medium Vegetation	16	Main power line	Power lines
6	Human structures	—	17	Other power line	Power lines
7	Low point (noise)	Noise	18	Fiber optic cable	—
8	Model keypoints (masspoints)	—	19	Not rated (object to be considered)	Noise
9	Water	—	20	Not rated (object to be ignored)	—
10	Rail	Ground	21	Incidents	—

Table 4. Distribution of semantic classes in the TS40K dataset.

Label	Semantic Class	Density (%)
0	Noise	1.348
1	Ground	55.281
2	Low Vegetation	35.520
3	Medium Vegetation	6.647
4	Power Line Support Tower	0.431
5	Power Line	0.771

labeled from a set of 22 classes detailed in Table 2.

Point Cloud Annotations by Maintenance Personnel. The annotations made by maintenance personnel are originally aimed to expedite power-grid inspections and not to train machine learning models. In collaboration with these experts, we mapped the maintenance classes to semantic categories that have clear contextual relevance in the 3D scenes (Tab. 3). These resulting classes represent critical scene elements that need to be identified during inspections to assess the risk of contact between the electrical system and its surroundings. This mapping process led to the class distribution illustrated in Tab. 5. Additionally, the topology of a power-grid system is sensitive information that could put in jeopardy both the system's safety and the security of the people who rely on it. To maintain accessibility while safeguarding this information, we partition the TS40K dataset into the three sample types and preprocess them (e.g., normalizing the coordinates in each sample) to prevent reverse engineering the original point clouds: **(1) Tower-radius**: Includes the environment around a power-line support tower, providing a comprehensive view of the surroundings relevant to the tower's location. **(2) Power-line**: Focuses on power lines as the main actors, featuring two towers at opposite sides. This sample type offers insights into the spatial relationships of power lines and their supporting structures. **(3) No-tower**: Represents rural terrain without supporting towers but potentially includes power lines. This sample type provides context for areas where transmission infrastructure is absent. On average, each sample type has a length of 70, 100, and 90 meters, respectively.

ML in Power-grid Inspections. To the best of our knowledge, TS40K is the first 3D dataset that focuses on power-grid inspections. In settings where UAVs are employed to retrieve 3D representations of a power-grid systems, our dataset can aid in significantly speed up inspection time and, thus, be crucial in preventing damages, power outages and forest fires. The meticulous process of annotating 3D points with inspection-based labels is extremely time-consuming and puts inspection efficacy at risk when considering the vast extent of power grid systems. To this end, developing machine learning tools to assist the labeling process enables faster inspections and safeguards power-grid systems.

Class Distribution Analysis. Table 4 illustrates the distribution of various classes. Notably, the *ground* and *low vegetation* classes are the most prevalent, while the classes that compose the electrical transmission system, i.e., *power line support tower* and *power line*, are under-represented. This imbalance is a common characteristic in datasets captured from a birds-eye view perspective and in natural environments such as rural areas. Other 3D benchmarks may also display class imbalance, although not to the same extent as ours. For example, SemanticKITTI [2] shows a lower point frequency in *human* classes, such as cyclist. In contrast, SensatUrban [20], S3DIS [1], and ShapeNet [4] demonstrate a balanced class distribution. Thus, our TS40K dataset presents a new challenge for model training, as some classes naturally occur less frequently. Addressing this unbalance becomes a crucial aspect for any effective approach aiming to handle diverse real-world scenarios.

4. Tasks and Benchmarks

Train/Test split statistics. The TS40K dataset has a total of 24,355 samples across its three sample types: 3663 in tower-radius, 3590 in power-line and 17,102 in no-tower. For each type, 80% of the samples were reserved for model

fitting and the remaining 20% for testing. The splitting between training and validation is done at random at each training cycle.

Sub-sampling Techniques. To slightly alleviate the class imbalance between the electrical transmission system and its environment while preserving the geometry of 3D scenes, we utilize farthest point sampling (FPS) [32]. FPS is a commonly employed technique in 3D scene understanding applications due to its ability to preserve the geometry of 3D elements and adjust class representation. However, its drawback lies in its time complexity. Alternative sub-sampling methods, such as inverse density importance sub-sampling (IDISS) [17] and random point sampling (RPS) [21], have gained attention for their more favorable time complexities. However, IDISS does not ensure the preservation of 3D scene geometry, as it selects K points with lower density within a ball of radius l. This not only prioritizes classes with lower density, such as noise, at the expense of others, but it may also lead to varying point densities for the same type of object across different 3D scenes, such as in lower vegetation. On the other hand, while RPS offers time efficiency, it may eliminate many points from underrepresented classes, compromising the accuracy of object segmentation. This trade-off is unacceptable in our case, where precise detection of the power-grid is paramount for ensuring its safe inspection. We demonstrate the pros and cons of these sub-sampling techniques in Fig. 4. As a result of employing FPS, the total density for the power-grid reaches 2.9%, representing a 1.7% increase compared to the original data distribution shown in Table 4.

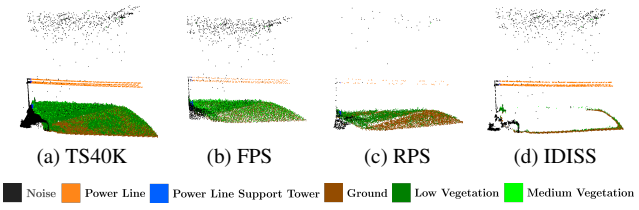


Figure 4. Comparison of three subsampling techniques for 3D point clouds. Figs. 4b to 4d depict the effects of Farthest Point Sampling (FPS), Random Point Sampling (RPS), and Inverse Density Importance Subsampling (IDISS), respectively. FPS preserves geometry and adjusts class representation, leading to a balanced distribution with increased density for the power-grid. IDISS prioritizes lower density classes, compromising the geometry preservation and causing inconsistent point densities. RPS, though time-efficient, may eliminate points from underrepresented classes, impacting object segmentation accuracy.

Table 5. **3D semantic segmentation**: Benchmark results of baselines on the TS40K test set. We report mean IoU (mIoU %) and per-class IoU (%) scores. Due to the extreme class imbalance, we showcase the results with both regular and weighted cross entropy.

Method	Loss Function	mIoU (%)	Ground	Low Vegetation	Medium Vegetation	Power Line Support Tower	Power Line
PointNet [40]	Cross Entropy	36.25	49.57	55.53	9.58	4.52	66.73
PointNet++ [41]		40.72	59.05	55.62	11.42	2.92	74.58
RandLaNet [21]		14.38	28.69	43.18	0.04	0	0
KPConv [54]		56.18	63.35	59.76	24.41	**40.62**	**92.75**
PTV1 [8, 70]		59.26	75.15	**66.02**	29.74	35.32	90.05
PTV2 [8, 60]		**62.27**	**77.73**	65.78	**49.45**	26.39	91.98
PointNet [40]	Weighted Cross Entropy	44.58	62.72	44.92	17.91	17.57	79.79
PointNet++ [41]		46.90	59.03	55.35	18.57	21.32	80.22
RandLaNet [21]		16.76	23.21	40.27	17.38	0.91	2.02
KPConv [54]		57.58	64.52	59.23	38.08	33.03	93.06
PTV1 [8, 70]		62.67	**77.34**	**67.90**	32.78	**43.80**	91.51
PTV2 [8, 60]		**65.58**	77.31	64.22	**48.94**	43.42	**93.99**

4.1. 3D Semantic Segmentation

4.1.1 Task and Metrics

For a comprehensive evaluation of model performance in 3D semantic segmentation, we rely on the mean intersection-over-union (mIoU) metric [13] across all classes:

$$\text{mIoU} = \frac{1}{C}\sum_{c=1}^{C}\frac{TP_c}{TP_c + FP_c + FN_c}. \quad (1)$$

Here, TP_c, FP_c, and FN_c stand for true positive, false positive, and false negative predictions for class c, and C is the total number of classes. The mIoU serves as a key metric for assessing the segmentation accuracy of 3D point cloud models. It quantifies the degree of overlap between predicted and ground truth segmentation masks.

4.1.2 Results and Discussion.

Standard Cross Entropy. Focusing on the first six rows of Tab. 5, Point Transformer V2 (PTV2) [60] achieves the highest mIoU of 62.27%. It demonstrates significant improvements in segments like Medium Vegetation (49.45%) when compared to Point Transformer (PTV1) [70] and KPConv [54]. Interestingly, KPConv achieves the highest IoU performance in power-grid segments, namely in power lines (92.75%) and their supporting towers (40.62%). This may be due to the way PTV2 and KPConv extract features: PTV2 divides the 3D scene in non-overlapping windows and applies a point-attention mechanism, whereas KPConv detects signals with convolutional kernels that often overlap to build a fine-grained feature descriptor of the input. For segments with a small number of points, a more detailed description of their neighborhood probably yields better results. In contrast, RandLaNet [21] performs the worst with a mean IoU of 14.38%, failing to detect any elements of the power-grid. The use of random point sampling in RandLaNet compromises performance in segments with a low point count (Fig. 4).

Weighted Cross Entropy. The last six rows of Tab. 5 show that weighted cross entropy loss improves the perfor-

mance of the semantic segmentation models in the TS40K dataset. PTV2 [60] maintains its position as the top-performing baseline, achieving a mIoU of 65.58%. With the weighted loss, transformer-based methods [60, 70] achieve a better segmentation of power-grid elements than KP-Conv [54]. Most likely because the point-transformer mechanism now pay special attention to detecting power-grid segments, whereas KPConv is robust to such weighting schemes. RandLaNet [21] continues to underperform when compared to other models with a mIoU of 16.76%.

Application in Power-grid Inspections. These 3D semantic segmentation baselines do not exhibit performance high enough to be employed in the task of power-grid maintenance. Specifically, the highest IoU recorded for power line supporting towers is 43.80%, which is considered too low by industry experts involved in this work. These experts have defined a performance threshold of 85% IoU for power-grid elements before machine learning models can be deployed in inspection pipelines. For power lines, state-of-the-art methods achieve an IoU of nearly 94% , though accurate detection is still hindered by noise. As shown in Fig. 7, noise hinders the detection of both power lines and their supporting towers, despite high overall performance. Noise remains a significant issue, particularly during inspections under suboptimal weather conditions, where noisy point clouds make accurate segmentation more challenging.

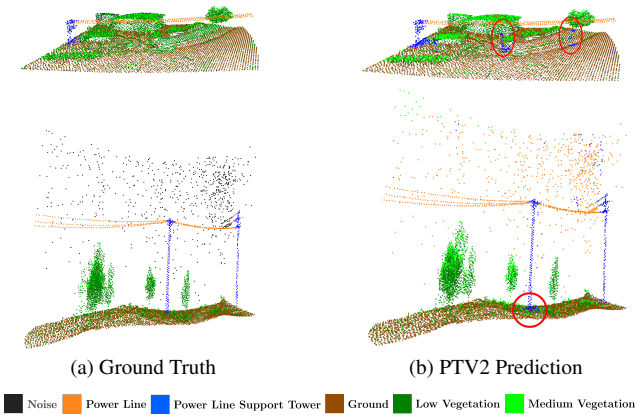
Qualitative Performance Analysis of Point Transformer V2 on TS40K. Figure 5 showcases the impact of noisy labels in the PTV2 [60] model. In the first row, PTV2 demonstrates robustness by accurately predicting the main tower while also detecting smaller voltage towers often missed in ground truth annotations. However, our analysis reveals instances, as seen in the second row of the same Figure, where PTV2 introduces artifacts such as additional ground patches not present in the original labels. This phenomenon highlights the challenge posed by noisy labels in TS40K, potentially affecting the reliability of model evaluations.

4.2. 3D Object Detection

4.2.1 Task and Metrics.

In 3D object detection, we detect three scene elements to assess the proximity of the power grid and vegetation. Specifically, we detect power lines, supporting towers and medium vegetation. To evaluate this, Average Precision (AP) is the metric of choice. It determines the area under the precision-recall curve and averages across all classes:

$$AP_c = \int_0^1 P_c(R)\, dR \quad \text{and} \quad mAP = \frac{1}{C}\sum_{c=1}^{C} AP_c. \quad (2)$$



(a) Ground Truth (b) PTV2 Prediction

Figure 5. Qualitative results showcasing the performance of Point Transformer V2 (PTV2) [60] on the TS40K dataset. In the first row, PTV2 successfully predicts the primary tower in the scene and accurately identifies smaller voltage towers, often overlooked in the ground truth annotations. However, the second row reveals an instance where PTV2 introduces a patch of ground surrounding two towers that was absent in the original labels. This highlights the impact of noisy labels in 3D benchmarks like PTV2.

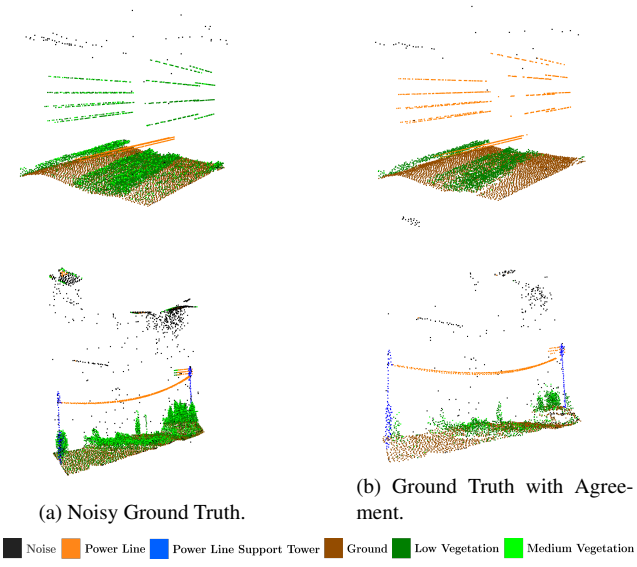
Table 6. **3D object detection**: Benchmark results of baselines on the TS40K test set under the 3D Average Precision (AP) metric with 11 recall points. We report mean AP and per-class AP scores.

Method	mAP	Power Line Support Tower	Power Line	Medium Vegetation
SECOND [63]	52.68	32.64	85.09	40.32
PointPillar [28]	56.63	38.63	83.74	47.53
Point-RCNN [45]	57.65	36.54	88.71	44.26
Part-A^2 net [46]	58.65	39.55	86.69	48.00
PV-RCNN [43]	**61.23**	**40.32**	**92.77**	**50.61**

Here, $P_c(R)$ denotes the precision at recall level R for class c. The integration spans the entire recall range, capturing the precision nuances. The resulting AP_c offers a granular evaluation of the model's effectiveness in discerning class-specific objects within the 3D space. This metric provides a detailed assessment of precision in 3D point cloud analysis and sheds light on the model's accuracy in classifying diverse structures within the point cloud data.

4.2.2 Results and Discussion.

Table 6 showcases the performance of leading 3D object detection baselines within the framework of OpenPCDet [53]. PV-RCNN [43] is the top-performing method with a mean AP of 61.23%, demonstrating superior performance across all evaluated categories. Other baselines like Point-RCNN, Part-A^2 net, and PointPillar also exhibit competitive mean AP scores. Other baselines like Point-RCNN, Part-A^2 net, and PointPillar also show competitive mean AP scores. While all methods perform well in detecting power lines, there is a notable discrepancy in detecting power line sup-



(a) Noisy Ground Truth.

(b) Ground Truth with Agreement.

Figure 6. Noisy labeling can be a significant challenge. In the TS40K dataset, instances of mislabeled 3D elements are apparent. For instance, occasional noise and power lines might be mistakenly classified as medium vegetation. This mislabeling can occur due to safety considerations around tower areas and the presence of power lines not connected to the main grid, which may not be properly identified. To mitigate this, we take advantage of SOTA models to select 3D points based on model agreement showed in the bottom row. This way, we can achieve a ground truth that is more consistent for model training.

porting towers. This may be due to the characteristics of the TS40K scenes, where power lines are isolated while supporting towers are often surrounded by other objects. The labeling process also introduces complexity, as regions around towers are sometimes mislabeled due to the top-down annotation convention used by maintenance personnel, leading to noisy labeling. For example, the ground and noise are occasionally mislabeled as towers (Fig. 6).

5. TS40K Unique Challenges

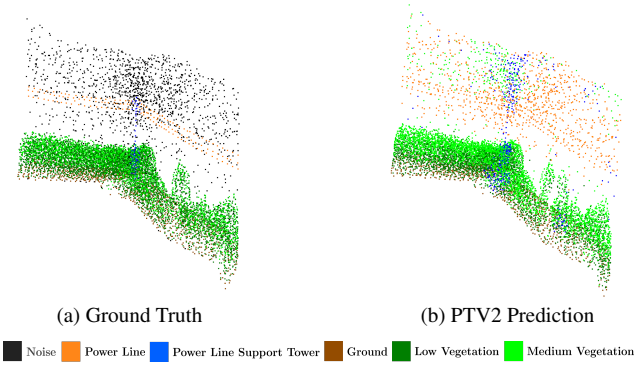
Noisy Labels. The acquisition and labeling process of TS40K raw 3D point clouds is a long and meticulous process that focuses on safeguarding the transmission system. The dataset's high point-density and homogeneous object density provide a detailed representation of rural environments, which coupled with the thousands of kilometers of power-grid to be closely inspected, leads to labeling mishaps when mapping the maintenance classes to semantically meaningful labels. Specifically, some points around supporting towers might be misclassified: Patches of ground below towers, some power line elements and occasional noise above them are, sometimes, mistakenly labeled as part of towers. In addition, a number of power lines may be misclassified as medium vegetation because they are not part of the main power-grid system. Thus, this mislabeling complicates 3D scene understanding evaluation (as illustrated in Fig. 5). To mitigate this, we take advantage of the agreement of state-of-the-art models to achieve a ground truth where noisy labels have less impact and, thus, is better suited to train and evaluate ML models. Specifically, we calculate the model agreement of the top-5 performing models on the TS40K dataset and select 3D points where the agreement is above 70%, but also keep all points labeled as supporting towers in the noisy ground truth seeing as models record a low performance for this class (Fig. 6). On the other hand, this significant pruning reduces the number of points in the dataset by 40%, so we provide the agreement scores and majority classes alongside the original version of the dataset plus this alternative version to assist researchers in making informed decisions about the data. Nevertheless, this noisy labeling highlights the obstacles inherent to using datasets not specifically curated for machine learning, and the inclusion of diverse scenarios ensures a broad coverage of real-world challenges. For the next steps, we will introduce a post-reviewing stage that employs a hybrid approach: a model's prediction is accepted if the softmax scores meet a certain threshold; otherwise, a human evaluator will make the final assessment.

Spurious Points from High-density Noise. Unmanned LiDAR drones capture power-grid environments but often encounter noise, particularly influenced by weather conditions. These spurious points can obscure power-grid elements, making accurate detection and scene interpretation more challenging. Figure 7 demonstrates the issue of high noise density in the TS40K dataset. While the noise class is typically disregarded during training for 3D semantic segmentation tasks due to its low relevance and unpredictable distribution, certain dataset samples contain a significant amount of noise. This noise impairs the accuracy of predictions in 3D models such as PTV2. As shown in Figure 7, the segmentation of towers and power lines becomes obscured by the noise, hindering the retrieval of power-grid elements. To gain deeper insights into the suboptimal performance of state-of-the-art models in detecting power-grid elements, we analyze the confusion matrix of the top-3 performing models (with the best model, PTV2 [60], presented in Tab. 7). The results reveal that, on average, 11% of noise points are mistakenly classified as towers and 42% as power lines. This shows that addressing real-world sensor noise offers a valuable opportunity for evaluating denoising techniques and to increase performance in the TS40K dataset.

Diverse Structures and the Impact of Extreme Class Imbalance. The TS40K dataset, like other real-world 3D benchmarks, presents a significant challenge due to extreme

Table 7. **Analyzing the confusion matrix of PTV2** [60]: Noise is frequently mislabeled as support towers or power lines, likely due to proximity. Medium and low vegetation are also occasionally misclassified as tower. This real-world sensor noise provides an opportunity to evaluate denoising methods.

	Noise	Ground	Low Veg.	Med. Veg.	Tower	Power Line
Noise	0.0000	0.0905	0.1435	0.2095	**0.1114**	**0.4452**
Ground	0.0000	0.8645	0.0653	0.0639	0.0055	0.0007
Low Veg.	0.0000	0.0670	0.7320	0.1812	0.0177	0.0020
Med. Veg.	0.0000	0.0189	0.1406	0.8170	**0.0192**	0.0052
Tower	0.0000	0.0016	0.0063	0.0065	0.9784	0.0072
Power Line	0.0000	0.0001	0.0019	0.0054	0.0180	0.9736



(a) Ground Truth (b) PTV2 Prediction

Figure 7. High noise-density challenge in Point Transformer V2 (PTV2) [60] on the TS40K dataset. While the noise class is typically disregarded during training for 3D semantic segmentation tasks, certain dataset samples contain a high number of noise 3D points. In this figure, the segmentation of towers and power lines is obscured by the noise, hindering the retrieval of the power-grid..

class imbalance, particularly in power-grid elements such as power lines and their supporting towers. This imbalance impacts the performance of the 3D scene understanding baselines assessed on our dataset. In both 3D semantic segmentation and 3D object detection (refer to Tables 5 and 6), the evaluated baselines demonstrate consistent performance trends. While all models detect power lines efficiently, their performance with respect to supporting towers consistently lags behind, despite similar class densities in the dataset. The diversity of the structures, particularly towers, contributes to this challenge. To address the imbalance, we trained state-of-the-art models with tower-inclusive samples (Tab. 8), which improved tower IoU but decreased vegetation IoU when tested on tower-inclusive data. However, when evaluating the entire test set, all methods showed decreased tower IoU, indicating overfitting. We also explored oversampling towers using SMOTE [5] and undersampling the ground using IDISS [17] (Fig. 4). The results in Tab. 9 indicate that class imbalance alone does not fully explain the poor tower segmentation: while SMOTE showed similar results to Tab. 5, IDISS worsened performance.

Table 8. **Focusing on towers**: Benchmark results of 3D semantic segmentation baselines on the TS40K trained with *tower-radius* and *power-line* sets.

Method	Test Set	mIoU (%)	Ground	Low Vegetation	Medium Vegetation	Power Line Support Tower	Power Line
PointNet [40]	Only Tower Including Test Set	34.35	50.38	52.07	0.9	7.15	61.16
PointNet++ [41]		39.55	48.47	50.54	2.37	30.06	66.51
RandLaNet [21]		4.28	5.26	0	16.12	0	0
KPConv [54]		47.46	48.40	26.82	30.85	42.69	88.53
PTV1 [8, 70]		50.29	64.12	25.98	35.46	41.71	78.10
PTV2 [8, 60]		**60.88**	75.08	**48.74**	43.24	**48.47**	**88.86**
PointNet [40]	Entire Test Set	32.77	48.63	50.51	2.18	4.73	57.82
PointNet++ [41]		34.78	45.17	50.60	1.03	16.56	60.51
RandLaNet [21]		4.13	2.07	0	18.58	0	0
KPConv [54]		42.38	49.40	18.54	29.10	**32.08**	82.77
PTV1 [8, 70]		44.89	69.47	28.53	31.90	16.43	78.10
PTV2 [8, 60]		**56.41**	77.35	55.06	40.98	25.79	**82.87**

Table 9. **Mitigating class imbalance**: Benchmark results of 3D semantic segmentation baselines on the TS40K trained with SMOTE and IDISS preprocess.

Method	Imbalance Technique	mIoU (%)	Ground	Low Vegetation	Medium Vegetation	Power Line Support Tower	Power Line
PointNet [40]	Oversampling: SMOTE	42.59	57.65	54.61	14.93	12.83	72.95
PointNet++ [41]		44.31	64.20	57.64	14.52	12.30	72.86
RandLaNet [21]		8.09	23.37	0	17.10	0	0
KPConv [54]		48.92	65.11	40.27	35.51	14.26	89.27
PTV1		59.65	73.52	52.30	40.97	40.16	91.32
PTV2 [8, 60]		**65.17**	79.42	62.87	47.41	**43.22**	**92.94**
PointNet [40]	Undersampling: IDIS	23.51	48.81	37.46	0.70	0.04	30.51
PointNet++ [41]		30.59	55.08	53.69	10.76	1.81	31.59
RandLaNet [21]		20.86	47.06	36.77	9.86	0.00	10.62
KPConv [54]		38.13	51.08	49.15	13.22	6.30	70.92
PTV1 [8, 70]		47.12	66.80	**58.24**	21.25	15.99	73.29
PTV2 [8, 60]		**49.80**	66.85	53.22	**30.35**	**18.66**	**79.91**

6. Conclusions

In this paper, we introduced the TS40K dataset, a novel contribution to 3D scene understanding focused on rural electrical transmission systems. This dataset addresses a critical gap in current research, offering challenges distinct from those found in existing urban datasets.

Our evaluation highlights that state-of-the-art methods are not yet sufficient for reliable use in power-grid inspections, particularly due to the poor detection of supporting towers and the misclassification of noise as power lines. Key challenges identified include the need for improved adaptation of real-world data labels, addressing extreme class imbalance, and managing high-density noise.

TS40K provides a valuable resource for research in 3D scene understanding and infrastructure inspection, supporting the development of more robust and accurate systems for real-world applications.

Acknowledgements

This research was funded by DM 351-2022 of the Italian Ministry of University and Research and with partial support from the Italian INDAM – GNAMPA group, by FCT I.P., through the strategic project NOVA LINCS (UIDB/04516/2020). It was also funded by "Sustainable Stone by Portugal" Nº C644943391-00000051/40 co-funded by Recovery and Resilience Plan and NextGeneration EU, and via the oc1-2024-TES-01 issued and implemented by the ENFIELD project, under the grant agreement No 101120657. Views expressed are those of the author(s) only and do not reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] Iro Armeni, Sasha Sax, Amir R Zamir, and Silvio Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*, 2017. 1, 2, 4

[2] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. In *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*, 2019. 2, 4

[3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 2

[4] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 2, 4

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority oversampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002. 8

[6] Xiaobai Chen, Aleksey Golovinskiy, and Thomas Funkhouser. A benchmark for 3d mesh segmentation. *Acm transactions on graphics (tog)*, 28(3):1–12, 2009. 2

[7] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3075–3084, 2019. 3

[8] Pointcept Contributors. Pointcept: A codebase for point cloud perception research. https://github.com/Pointcept/Pointcept, 2023. 5, 8

[9] Giovanni Curnis, Simone Fontana, and Domenico G Sorrenti. Gtasynth: 3d synthetic data of outdoor non-urban environments. *Data in Brief*, 43:108412, 2022. 2

[10] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 1, 2

[11] Angela Dai and Matthias Nießner. 3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 452–468, 2018. 3

[12] Martin Engelcke, Dushyant Rao, Dominic Zeng Wang, Chi Hay Tong, and Ingmar Posner. Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1355–1361. IEEE, 2017. 3

[13] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111:98–136, 2015. 5

[14] Nils Gählert, Nicolas Jourdan, Marius Cordts, Uwe Franke, and Joachim Denzler. Cityscapes 3d: Dataset and benchmark for 9 dof vehicle detection. *arXiv preprint arXiv:2006.07864*, 2020. 2

[15] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 2

[16] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012. 2

[17] Fabian Groh, Patrick Wieschollek, and Hendrik PA Lensch. Flex-convolution: Million-scale point-cloud learning beyond grid-worlds. In *Asian Conference on Computer Vision*, pages 105–122. Springer, 2018. 5, 8

[18] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(12):4338–4364, 2020. 3

[19] Yuenan Hou, Xinge Zhu, Yuexin Ma, Chen Change Loy, and Yikang Li. Point-to-voxel knowledge distillation for lidar semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8479–8488, 2022. 3

[20] Qingyong Hu, Bo Yang, Sheikh Khalid, Wen Xiao, Niki Trigoni, and Andrew Markham. Towards semantic segmentation of urban-scale 3d point clouds: A dataset, benchmarks and challenges. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4977–4987, 2021. 2, 4

[21] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11108–11117, 2020. 3, 5, 6, 8

[22] Binh-Son Hua, Quang-Hieu Pham, Duc Thanh Nguyen, Minh-Khoi Tran, Lap-Fai Yu, and Sai-Kit Yeung. Scenenn: A scene meshes dataset with annotations. In *2016 fourth international conference on 3D vision (3DV)*, pages 92–101. Ieee, 2016. 2

[23] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1037–1045, 2015. 2

[24] Maximilian Jaritz, Jiayuan Gu, and Hao Su. Multi-view pointnet for 3d scene understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 3

[25] R. Kesten, M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet. Lyft level 5 av dataset 2019. https://level5.lyft.com/dataset/, 2019. 2

[26] Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. Rethinking range view representation for lidar segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 228–240, 2023. 1, 3

[27] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. Spherical transformer for lidar-based 3d recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17545–17555, 2023. 3

[28] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12697–12705, 2019. 6

[29] Felix Järemo Lawin, Martin Danelljan, Patrik Tosteberg, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Deep projective 3d semantic segmentation. In *Computer Analysis of Images and Patterns: 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part I 17*, pages 95–107. Springer, 2017. 3

[30] Truc Le and Ye Duan. Pointgrid: A deep network for 3d shape understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9204–9214, 2018. 3

[31] Bo Li. 3d fully convolutional network for vehicle detection in point cloud. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1513–1518. IEEE, 2017. 3

[32] Jingtao Li, Jian Zhou, Yan Xiong, Xing Chen, and Chaitali Chakrabarti. An adjustable farthest point sampling method for approximately-sorted point cloud data. In *2022 IEEE Workshop on Signal Processing Systems (SiPS)*, pages 1–6. IEEE, 2022. 5

[33] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018. 3

[34] Yecheng Lyu, Xinming Huang, and Ziming Zhang. Learning to segment 3d point clouds in 2d image space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12255–12264, 2020. 3

[35] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, Jingheng Chen, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, et al. One million scenes for autonomous driving: Once dataset. *arXiv preprint arXiv:2106.11037*, 2021. 2

[36] Jiageng Mao, Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. 3d object detection for autonomous driving: A review and new outlooks. *arXiv preprint arXiv:2206.09474*, 2022. 1, 3

[37] Sergio Marconi, Sarah J Graves, Dihong Gong, Morteza Shahriari Nia, Marion Le Bras, Bonnie J Dorr, Peter Fontana, Justin Gearhart, Craig Greenberg, Dave J Harris, et al. A data science challenge for converting airborne remote sensing data into ecological information. *PeerJ*, 6:e5843, 2019. 2

[38] Hsien-Yu Meng, Lin Gao, Yu-Kun Lai, and Dinesh Manocha. Vv-net: Voxel vae net with group convolutions for point cloud segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8500–8508, 2019. 3

[39] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 909–918, 2019. 2

[40] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 3, 5, 8

[41] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 3, 5, 8

[42] B Don Russell, Carl L Benner, and Jeffrey A Wischkaemper. Distribution feeder caused wildfires: Mechanisms and prevention. In *2012 65th Annual Conference for Protective Relay Engineers*, pages 43–51. IEEE, 2012. 1

[43] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10529–10538, 2020. 6

[44] Shaoshuai Shi, Li Jiang, Jiajun Deng, Zhe Wang, Chaoxu Guo, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn++: Point-voxel feature set abstraction with local vector representation for 3d object detection. *International Journal of Computer Vision*, 131(2):531–551, 2023. 3

[45] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 770–779, 2019. 3, 6

[46] Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2647–2664, 2020. 6

[47] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V 12*, pages 746–760. Springer, 2012. 2

[48] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 567–576, 2015. 2

[49] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015. 3

[50] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 2

[51] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching efficient 3d architectures with sparse point-voxel convolution. In *European conference on computer vision*, pages 685–702. Springer, 2020. 3

[52] B Teague, R McLeod, and S Pascoe. Final report, 2009 victorian bushfires royal commission. *Parliament of Victoria, Melbourne Victoria, Australia*, 1, 2010. 1

[53] OpenPCDet Development Team. Openpcdet: An open-source toolbox for 3d object detection from point clouds. https://github.com/open-mmlab/OpenPCDet, 2020. 6

[54] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6411–6420, 2019. 3, 5, 6, 8

[55] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011. 1

[56] Jan Trochta, Martin Krucek, Tomas Vrska, and Kamil Kral. 3d forest: An application for descriptions of three-dimensional forest structures using terrestrial lidar. *PloS one*, 12(5):e0176871, 2017. 2

[57] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1588–1597, 2019. 2

[58] Nina Varney, Vijayan K Asari, and Quinn Graehling. Dales: A large-scale aerial lidar data set for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 186–187, 2020. 2

[59] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler, faster, stronger, 2023. 3

[60] Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, and Hengshuang Zhao. Point transformer v2: Grouped vector attention and partition-based pooling. In *NeurIPS*, 2022. 5, 6, 7, 8

[61] Qiangeng Xu, Yiqi Zhong, and Ulrich Neumann. Behind the curtain: Learning occluded shapes for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2893–2901, 2022. 3

[62] Xu Yan, Jiantao Gao, Jie Li, Ruimao Zhang, Zhen Li, Rui Huang, and Shuguang Cui. Sparse single sweep lidar point cloud segmentation via learning contextual shape priors from scene completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3101–3109, 2021. 1

[63] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018. 6

[64] Bin Yang, Ming Liang, and Raquel Urtasun. Hdnet: Exploiting hd maps for 3d object detection. In *Conference on Robot Learning*, pages 146–155. PMLR, 2018. 3

[65] Zetong Yang, Yanan Sun, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Ipod: Intensive point-based object detector for point cloud. *arXiv preprint arXiv:1812.05276*, 2018. 3

[66] Ze Yang and Liwei Wang. Learning relationships for multi-view 3d object recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7505–7514, 2019. 3

[67] Jesus Zarzar, Silvio Giancola, and Bernard Ghanem. Pointrgcn: Graph convolution networks for 3d vehicles detection refinement. *arXiv preprint arXiv:1911.12236*, 2019. 3

[68] Yifan Zhang, Qijian Zhang, Zhiyu Zhu, Junhui Hou, and Yixuan Yuan. Glenet: Boosting 3d object detectors with generative label uncertainty estimation. *International Journal of Computer Vision*, 131(12):3332–3352, 2023. 3

[69] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9601–9610, 2020. 3

[70] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021. 5, 6, 8

[71] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4490–4499, 2018. 3